Recognition of objects in infrared band

Daniil Loktev^{1,*}

¹ Bauman Moscow State Technical University (National Research University), 2-nd Baumanskaya, 5, 105005, Moscow, Russia

Abstract. Within the course of the present research, the mathematical and algorithmic supporting of hardware-software system of control and monitoring that uses images obtained in visible and infrared wavelength band as the source of initial data has been developed and implemented. The paper defines the conditions for obtaining an image, the quality of which is sufficient to detect a target object. The front-end of software package has been worked out that allows an operator to address tasks on detecting, tracking and identifying objects without specific knowledge and experience.

1 Introduction

Modern systems of transport safety, monitoring and control of passenger flows are associated with detecting, capturing and recognizing great amounts of moving objects, including people's faces, as well as determining particular parameters of their condition and behavior through analyzing large arrays of photo and video information obtained in real time. One of the promising directions for the development of such systems is the identification of individual persons by their biometric personal data. Thereat, not only determining person's personality is important, but also estimating his or her psychophysical state, behavior adequacy, forecasting and tracking his or her routes at transport infrastructure facilities or rolling stocks. The climate pattern of Russia is characterized by low temperatures prevailing during the most part of a year, so passengers wear hats, scarves and other pieces of clothes that reduce the visibility of face and consequently its availability for being used as an image of a recognizable object. Glasses of various types and forms are also very popular, which complicates the process of recognizing a person's face as well. Such conditions demand using visual images obtained not only in the visible wavelength band, but also in the infrared band, since such images contain additional information on an object under study and can ensure sufficiently accurate detection and identification of various objects that are difficult to observe with the use of traditional video and photo detectors.

* Corresponding author: loktevdan@yandex.ru

2 Methods of research

The present study proposes to use a thermographic camera with sensitivity of 0.05 °C in order to obtain images of objects under study. Since objects being detected and recognized are mobile, although they move at relatively low speeds, there is a need to have several approaches for solving the problem. The author proposes to consider in detail two of them. When detecting and capturing the studied objects, it is proposed to divide them into groups “far” and “close” according to the distance between the objects and thermographic camera, and into groups “fast” and “slow” according to their speed. For the primary classification of objects according to their distance, their speed and type, the method of image blur analysis is preferable.

Setting the boundaries of these groups of objects is performed with the use of statistical analysis tools and based on fuzzy logic.

Within the next approximation, an object can be recognized by other features characterizing it (for instance, by color components).

As far as the presented monitoring system examines images obtained in the infrared band, the considered approaches and algorithms are not suitable for “fast” objects, since heat transfer is a rather slow process. The transfer of heat in space obeys the equation:

$$\Delta Q = -k \frac{\partial T}{\partial n} \Delta S, \quad (1)$$

where ΔQ is the heat amount, ΔS is the area of a certain surface through which the heat under consideration transfers, T is the temperature in the position with spatial coordinates (x, y, z) at a certain point in time t , n is the unit normal to the surface in the direction of heat transfer, k is thermal conductivity coefficient of the medium, where the monitoring system operates. The medium is considered to be homogeneous and isotropic.

The derivative with respect to the normal unit vector can be represented as:

$$\frac{\partial T}{\partial n} = \frac{\partial T}{\partial x} \cos \alpha + \frac{\partial T}{\partial y} \cos \beta + \frac{\partial T}{\partial z} \cos \gamma, \quad (2)$$

where $\cos \alpha, \cos \beta, \cos \gamma$ are directing cosines of a normal unit vector

In the general case, the heat-transfer equation can take the following form:

$$\frac{\partial T}{\partial t} = \frac{k}{c\rho} \Delta T, \quad (3)$$

where c and ρ are heat capacity and density of the substance, respectively,

$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is Laplace operator.

To solve equation (3), it is necessary to set the boundary and initial conditions and select an expression for a target temperature function, the coefficients of which will determine the problem specification in a particular medium.

Since passengers' movement through transport infrastructure is associated with redistribution of heat fluxes due to open doors, drafts and other processes that occur in public buildings with high traffic, boundary and initial conditions for thermoelastic phenomena will often change, which will also affect the time of equipment calibration when receiving the initial image of an object in the infrared band. The target functions included in defining equations can be expressed as an expansion in time and spatial coordinate that determines the studied point of the medium and the location of the wave front of value change

$$Z(s, t) = \sum_{k=0}^{\infty} \frac{1}{k!} [Z_{(k)}]_{t=s/G} \left(t - \frac{s}{G}\right)^k H\left(t - \frac{s}{G}\right), \quad (4)$$

where $Z_{(k)} = \partial^k Z / \partial t^k$ is the leap of time derivative of k -th order of the target function on the wave surface, $H(t-s/G)$ is the Heaviside unit function that determines the value leap at the front, s is the length of the arc measured in the direction of wave surface spread, t is the time counted from the start of heat exposure.

The proposed expansion of the unknown function relates to the ray method, which also applies geometric and kinematic compatibility conditions

$$G \left[\frac{\partial z_{(k)}}{\partial s} \right] = -[Z_{s(k+1)}] + \frac{\varepsilon[z_{(k)}]}{\delta t}, \quad (5)$$

where $\frac{\varepsilon}{\delta t} - \delta$ is the time derivative at the front of the wave surface, the spatial coordinates used in this expression are simultaneously the surface coordinates at the front of the wave surface, all three coordinate lines satisfy the conditions of mutual orthogonality, the time that is in denominator is actually determined relative to the wave front.

This article does not consider the task of assessing the movement speed of identification object relative to the speed of propagation of the heat flux and the speed of acquiescing image in the infrared band by thermographic camera. Therefore, the speed of movement of a passenger flow and individuals is considered quite slow, objects themselves are assumed located close to the photodetector. These assumptions will enable not taking into account the inertia of ongoing thermal processes in subsequent algorithms and reducing the task of visualizing an object in the infrared band to obtaining an image with a thermographic camera for wavelengths from 6 to 20 μm , which corresponds to heat radiated by human body.

For direct object recognition, it is proposed to use two methods based on the construction of cascading classifiers based on Haar features and background subtraction algorithms. Since the resulting images in the infrared band under consideration have fewer colors than a normal photo, the application of the described approaches seems reasonable.

When constructing cascading classifiers, the adaptive boosting algorithm (AdaBoost) [1] is used. This algorithm implies the synthesize of single complex and reliable classifier from a variety of simple ones. The process of synthesis and subsequent training of the final classifier focuses on “worse” recognizable elements.

At each stage, AdaBoost calls the most unreliable classifier and redistributes weight coefficients D_t characterizing the importance of each object from the training set. Each iteration of the weight distribution leads to the fact that the weight coefficient increases for an incorrectly classified object and the next classifier “focuses” on these objects, minimizing the weighted deviation:

$$h_t = \arg \min_{h_j \in H} \varepsilon_j, \text{ where } \varepsilon_t = \sum_{i=1}^m D_t(i) [y_i \neq h_j(x_i)], \quad (6)$$

ε_t is weighted deviation of a certain classifier h_t , when $\varepsilon_t \geq 0.5$ the execution of this algorithm stops.

The final classifier can be represented as

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right), \quad (7)$$

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\varepsilon_t}.$$

here the agreed designation is

Weight coefficients D_t determined at any stage of the procedure should be determined so that the following condition is satisfied:

$$\begin{cases} e^{-\alpha_t y_i h_t(x_i)} < 1, y(i) = h_t(x_i) \\ e^{-\alpha_t y_i h_t(x_i)} > 1, y(i) \neq h_t(x_i) \end{cases} \quad (8)$$

After completing the procedure for selecting the optimal classifier h_t for the distribution of weighting coefficients D_t , objects x_i that are recognized by the corresponding classifier h_t reliably have final weights smaller than objects that are unreliably identified. When testing classifiers for the distribution of next-order weights D_{t+1} , the algorithm selects a classifier, which better identifies objects that are incorrectly recognized during the previous step. To construct individual classifiers, Haar features [2] are used, working according to the

following scheme: any of the simplest features is imposed on the reference image, and then the difference between sums of values of white and black pixels is calculated. As a result, the value of generalized anisotropy characteristic of the selected section N of an image is obtained

$$N = Q_w - Q_{bl}, \quad (9)$$

where Q_w, Q_{bl} are total numbers of white and black pixels in features superimposed on the selected area of an image.

The implementation of the presented procedure requires significant computing power, since even for a small image the number of simple primitives (features) is large. The main characteristic of the modified adaptive amplification algorithm is that the features describing this object most reliably are selected.

The second approach used in this study is the method of subtracting the background, therewith more than one model (several Gaussian models besides average value and variance) are used to represent each pixel. It is assumed that each pixel has 3 Gaussian models. If the pixels do not match the Gaussian background distribution, then they are considered to be in the foreground. The available background subtraction techniques differ significantly from each other, but all of them imply that the observed series of images I consists of a static background B and objects moving in front of it [3]. Any moving object is assumed to have a color distribution that is different from the background. In general case, background subtraction methods can be represented as a ratio:

$$\chi_t(s) = \begin{cases} 1, & \text{when } d(I_{s,t}, B_s) > \tau, \\ 0 & \text{or} \end{cases} \quad (10)$$

here χ_t is motion mask at time point t , $I_{s,t}$ - color of pixel s at time point t , B_s - background of pixel s , d - distance between $I_{s,t}$ and B_s , τ - some threshold value.

Several ways to detect motion in an image are considered further. The simplest way to obtain background B is to create a single gray or color image that does not contain moving objects. For this purpose, a transport infrastructure facility is captured without moving objects (people) or a picture is taken using a median filter [4]. To reduce the influence of illumination and changes, the background can be expressed through the following iterative expression:

$$B_{s,t+1} = (1 - \alpha)B_{s,t} + \alpha I_{s,t}, \quad (11)$$

where $\alpha = 0..1$ is some certain constant.

The presented background model allows defining the pixels belonging to moving objects located in the foreground by determining the threshold of distance functions of different orders:

$$d_0 = |I_{s,t} - B_{s,t}|, \quad (12)$$

$$d_1 = |I_{s,t}^R - B_{s,t}^R| + |I_{s,t}^G - B_{s,t}^G| + |I_{s,t}^B - B_{s,t}^B|, \quad (13)$$

$$d_2 = (I_{s,t}^R - B_{s,t}^R)^2 + (I_{s,t}^G - B_{s,t}^G)^2 + (I_{s,t}^B - B_{s,t}^B)^2, \quad (14)$$

$$d_{\infty} = \max\{|I_{s,t}^R - B_{s,t}^R| + |I_{s,t}^G - B_{s,t}^G| + |I_{s,t}^B - B_{s,t}^B|\}, \quad (15)$$

where indices R, G and B denote the intensity of red, green and blue in the marked pixel, d_0 is the initial measure of distance, determined in shades of gray.

The proposed scheme allows using the previous frame I_{t-1} as the background image B , this approach enables reducing the final computational complexity of the procedure. This approach allows makes possible detecting movement by comparing neighboring frames, it is resistant to illumination changing in the whole picture, however, it is difficult to detect the entire moving object, but not its parts.

Pixels belonging to the background can be detected using *MinMax* method, which defines a condition, which is when satisfied is the criteria of a pixel being classified as a static object:

$$|M_s - I_{s,t}| < \tau d_\mu \quad (16)$$

where τ is the threshold specified by an operator of the monitoring system, d_μ is the median value of the largest absolute difference between frames throughout the image, s is the background pixel, M_s is the maximum difference between frames, which is associated with the minimum value of this difference m_s and the maximum difference of consecutive frames D_s observed during the series of images under study [5].

Condition (16) takes into account the fact that in a noisy area of the image, a pixel should have a larger change than in an area of a stable background. This approach implies that describing a target background pixel by the three extremum values m_s , M_s , D_s is more informative than using the traditional average vector and covariance matrix. Since condition (16) works with grayscale images, some of the information may be lost in comparison with color frame sequences.

The assessment of pixel's belonging to a background image can also be performed through modeling a multimodal probability distribution function [6, 7]:

$$P(I_{s,t}) = \frac{1}{N} \sum_{i=t-N}^{t-1} K(I_{s,t} - I_{s,i}), \quad (17)$$

here K is the kernel, N is the number of previous frames, $P(I_{s,t})$ is the estimated probability.

If the assessment is based on a sequence of color frames, then one-dimensional kernels can be used in expression (17):

$$P(I_{s,t}) = \frac{1}{N} \sum_{i=t-N}^{t-1} \prod_{j \in \{R,G,B\}} K\left(\frac{(I_{s,t}^j - I_{s,i}^j)}{\sigma_j}\right), \quad (18)$$

in this expression, σ_j can be definitive or preliminary estimated.

If the background image is structurally and chromatically complex, then the multimodal probability distribution functions can be used. In this case, each pixel is modeled by a set of K Gaussians, the probability of a certain color appearing in a given pixel s can be represented:

$$P(I_{s,t}) = \sum_{i=1}^K \omega_{i,s,t} N(\mu_{i,s,t}, \Sigma_{i,s,t}), \quad (19)$$

here $N(\mu_{i,s,t}, \Sigma_{i,s,t})$ – representation of the i -th model of the Gaussian, $\omega_{i,s,t}$ is the weight of the i -th model.

When calculating, it can be assumed that $\Sigma_{i,s,t} = \sigma^2 Id$, i.e. the covariance matrix is diagonal and the values engaged in (19) for a Gaussian, the intensity $I_{s,t}$ of which does not exceed a given deviation from the average value, are determined by the following recurrence relations:

$$\omega_{i,s,t} = (1 - \alpha) \omega_{i,s,t-1} + \alpha, \quad (20)$$

$$\mu_{i,s,t} = (1 - \rho) \mu_{i,s,t-1} + \rho I_{s,t}, \quad (21)$$

$$\sigma_{i,s,t}^2 = (1 - \rho) \sigma_{i,s,t-1}^2 + \rho d_2(I_{s,t}, \mu_{i,s,t}), \quad (22)$$

where α is the set value that shows learning speed of the algorithm, $\rho = \alpha N(\mu_{i,s,t}, \Sigma_{i,s,t})$, ρ is the second approximation of learning speed, d_2 is the distance between pixels.

It is assumed that the values μ and σ for mismatched distributions do not change, and only their weight changes in the direction of decreasing $\omega_{i,s,t} = (1 - \alpha) \omega_{i,s,t-1}$ [8, 9]. If at a certain stage of the iterative process the component under consideration does not correspond to the color $I_{s,t}$ with the smallest weight, then it is replaced by a Gaussian with a large initial dispersion σ_0^2 and a lower weight coefficient ω_0 . After overriding each Gaussian, the weighting factors are normalized and summed until the sum of a unit value is

reached. After that, K distributions are ordered according to relation $\frac{\omega_{i,x,t}}{\sigma_{i,x,t}}$ and H of the most reliable ones are determined as the background:

$$H = \operatorname{argmin}_h (\sum_{i=1}^h \omega_i > \tau). \quad (23)$$

After the procedure described above, pixels having color $I_{x,t}$, that differs by a larger deviation than the specified one from all the obtained distributions H , are detected as belonging to the moving object.

If there is significant noise in initial images and when using algorithms based on the Gaussian model, noise components can increase. In order to reduce this effect, morphological operators – erosion and dilation (expansion) – are proposed for the use.

The binary representation of erosion is:

$$A \ominus B = \bigcap_{b \in B} A_b, \quad (24)$$

where A is the main binary image, B is the binary representation of the structural element causing erosion. Image B passes through the whole image A and if single pixel A is similar to pixel B , then the logical addition of central pixel B with corresponding pixel A occurs, so the original image is cleared of objects smaller than the structural element.

The binary extension operator is represented as:

$$A \oplus B = \bigcup_{b \in B} A_b. \quad (25)$$

If the coordinate origin of the structural element B coincides with a single pixel, then the transfer and subsequent addition with corresponding pixels A is applied to the whole element B .

Operator (24) is mainly used to clear the background, and operator (25) is used to detect foreground objects. In fact, both erosion and dilation deal with object boundaries, primarily affecting small graphic elements, so often the detection of object boundaries and corner points is a separate task. One way to detect a boundary of an object is to use Canny operator.

The described operator is most often implemented for images in shades of gray in order to reduce the cost of computing power.

The outline defined this way is an array of points connected into a curve. Each foreground object (movable) is characterized by its own outline. This procedure is intended to help in detecting the overlap of studied objects and marking objects that have all points on the outline curve and further are to be recognized and classified [10, 11].

The face recognition system under development contains the following modules: loading a series of images of passenger flow, entering user recognition parameters, background subtraction algorithm, matching mode for subsequent frames, detecting an individual passenger, recognizing an object or recording into the database.

Automated system for monitoring and controlling passenger flows should have the following modules: receiving and loading a series of passenger images in the specified room, entering passenger detection parameters, starting the background subtraction algorithm, detecting and capturing a passenger image, recognizing and recording into the database [12-15].

The present study, when considering an integrated monitoring system, paid special attention to the unit of detecting an object itself that uses the background subtraction algorithm, which should manage to take into account: sudden or gradual change in illumination; repetitive or oscillatory movements of individual elements in the background; long-term changes in the position of objects in the big picture.

The background subtraction procedure is quite well known and is used in many image editors and image-processing programs to create a foreground mask (a binary image including pixels related to moving objects) in various modifications. The foreground mask is determined by subtracting the background image from the current frame, the background

image is formed taking into account the parameters of observed picture and objects peculiar to time of changes in the position of individual objects and photo detector settings.

Implementation of the presented algorithm is possible using OpenCV technical vision library, which also allows working with cascading classifiers based on the use of Haar features. The method of segmenting the background and foreground objects using a set of Gaussian for each element is used as a subtraction method, in the library this method is denoted as MOG2. The specified algorithm selects the most accurate Gaussian distribution for each pixel and therefore adapts well to changing shooting conditions.

Several functions are used sequentially to perform the described procedure, the first of which is `setShadowValue` which is responsible for detecting and indicating shadows and has one parameter that takes values from 0 to 255.

Then “`apply`” function is executed, which finds the foreground mask and has three parameters: “`image`”, “`fgmask`”, “`learningRate`”. “`Image`” defines the next frame of the sequence, which is used without scaling. “`Fgmask`” sets the foreground mask as an 8-bit binary image. “`LearningRate`” is the parameter that sets the speed of background change, when it is equal to zero, it means that the background remains the same for the entire series of images, and equaling one indicates that the background is redefined every time according to the last image in the series. This parameter can take negative values, in this case, the change speed of the background is automatically selected.

At the next stage, “`setHistory`” function is used, which determines the number of frames taken into account when receiving the model of the background.

To detect and select individual moving objects, it is proposed to use “`findAndDrawContours`” function, which detects outlines in a binary representation. If the found outlines are larger than the user-defined threshold value indicated by “`areaThreshold`” parameter, then objects referring to corresponding outlines are marked by rectangular frames with the use of “`drawBoundingBox`” function,.

In winter or other periods when people wear hats and other clothing, with excessive hair, as well as with the presence of various accessories, the image area of an object (passenger's face) can significantly decrease, which complicates the task of recognizing passengers' faces. To obtain additional information on the object, it is proposed to use not only the visible wavelength band, but also infrared band. Since thermal processes are sufficiently inertial, when receiving an image in the infrared band, it takes significantly more time for exposure and obtaining the initial image having quality that is sufficient to use algorithms of object detection and recognition.

3 Conclusions

When the mathematical and algorithmic support proposed in this study is implemented in the form of a software-hardware system for monitoring and control, it will allow detecting, tracking, and recognizing passengers in both visible and infrared wavelength bands, being based on the knowledge and competencies of a regular operator of an automated workstation equipped with a computer. The monitoring system will enable observing a big picture after the implementation of background subtraction algorithms, which will allow an operator to detect a target object in space, even when there is high density of passengers.

The presented approach of using the infrared wavelength band and the constitutive relations considered in the paper allow one to select the functioning parameters correctly for both monitoring system and thermographic camera recommended for solving the stated problem: infrared detector resolution, sensitivity, temperature range, thermographic camera resolution, measurement accuracy, wavelength range.

References

1. P. Gil, J. Pomares, F. Torres, *International Conference On Pattern Recognition*, P. 311-315 (2010).
2. T. Gaspar, P. Oliveira *Preprints of the 18th IFAC World Congress Milano (Italy)*. P. 9470-9475 (2011)
3. H. Hu, G. Haan *Proceedings IEEE International Conference on Image Processing, Atlanta*, P. 617-620 (2006)
4. T. Zaman Depth Estimation from Blur Estimation. <http://www.timzaman.com>, (2012).
5. H.-Y. Lin, C.-H. Chang *Optical Engineering* **45(12)**, N127201, 1-12. (2006)
6. D.A. Loktev, A.A. Loktev *Contemporary Engineering Sciences* **8**. n. 9. P. 467-475. (2015)
7. D.A. Loktev, A.A. Loktev *Contemporary Engineering Sciences*. **8**. n. 20. P. 789-797. (2015)
8. B. Glusberg, A. Loktev et al., *Advances in Intelligent Systems and Computing*, Vol **982**, pp 337-345 (2020)
9. A.A. Loktev, V.V. Korolev et al., *Procedia Engineering* **189**, 133 – 137 (2017)
10. B. Glusberg, V. Korolev et al., *MATEC Web of Conferences* **239**, 01054 (2018)
11. A.A. Loktev, V.V. Korolev, I.V. Shishkina, *IOP Conf. Series: Materials Science and Engineering* **463**, 032019 (2018)
12. A.A. Loktev, V.V. Korolev et al., *IOP Conf. Series: Materials Science and Engineering*, **463**, 032018 (2018)
13. B. Glusberg, A. Savin et al., *Advances in Intelligent Systems and Computing* **982**, 556-570 (2020)
14. A. Loktev, V. Korolev et al., *Advances in Intelligent Systems and Computing* **982**, 325-336
15. B. Glusberg, A. Savin et al., *Advances in Intelligent Systems and Computing* **982**, 571-587